AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology®

LETTER TO THE EDITOR

Orthervirales: New Virus Order Unifying Five Families of Reverse-Transcribing Viruses

Mart Krupovic,[a] Jonas Blomberg,[b] John M. Coffin,[c] Indranil Dasgupta,[d] Hung Fan,[e] Andrew D. Geering,[f] Robert Gifford,[g] Balázs Harrach,[h] Roger Hull,[i] Welkin Johnson,[j] Jan F. Kreuze,[k] Dirk Lindemann,[l] Carlos Llorens,[m] Ben Lockhart,[n] Jens Mayer,[o] Emmanuelle Muller,[p,q] Neil E. Olszewski,[r] Hanu R. Pappu,[s] Mikhail M. Pooggin,[t] Katja R. Richert-Pöggeler,[u] Sead Sabanadzovic,[v] Hélène Sanfaçon,[w] James E. Schoelz,[x] Susan Seal,[y] Livia Stavolone,[z,aa] Jonathan P. Stoye,[bb] Pierre-Yves Teycheney,[cc,dd] Michael Tristem,[ee] Eugene V. Koonin,[ff] Jens H. Kuhn[gg]

[a]Department of Microbiology, Institut Pasteur, Paris, France
[b]Department of Medical Sciences, Uppsala University, Uppsala, Sweden
[c]Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, Massachusetts, USA
[d]Department of Plant Molecular Biology, University of Delhi, New Delhi, India
[e]Department of Molecular Biology and Biochemistry, University of California, Irvine, California, USA
[f]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, Queensland, Australia
[g]MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom
[h]Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Budapest, Hungary
[i]Child Okeford, Blandford Forum, Dorset, United Kingdom
[j]Biology Department, Boston College, Chestnut Hill, Massachusetts, USA
[k]Crop and System Sciences Division, International Potato Center (CIP), Lima, Peru
[l]Institute of Virology, Technische Universität Dresden, Dresden, Germany
[m]Biotechvana, Parc Cientific, Universitat de Valencia, Valencia, Spain
[n]Department of Plant Pathology, University of Minnesota, St. Paul, Minnesota, USA
[o]Institute of Human Genetics, University of Saarland, Homburg, Germany
[p]CIRAD, UMR BGPI, Montpellier, France
[q]BGPI, Université Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France
[r]Department of Microbial and Plant Biology, University of Minnesota, St. Paul, Minnesota, USA
[s]Department of Plant Pathology, Washington State University, Pullman, Washington, USA
[t]INRA, UMR BGPI, Montpellier, France
[u]Julius Kühn-Institut, Institute for Epidemiology and Pathogen Diagnostics, Braunschweig, Germany
[v]Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, Mississippi, USA
[w]Agriculture and Agri-Food Canada, Summerland Research and Development Centre, Summerland, BC, Canada
[x]Division of Plant Sciences, University of Missouri, Columbia, Missouri, USA
[y]Natural Resources Institute, University of Greenwich, Chatham, Kent, United Kingdom
[z]Consiglio Nazionale delle Ricerche, Istituto per la Protezione Sostenibile delle Piante, Bari, Italy
[aa]International Institute of Tropical Agriculture, Ibadan, Nigeria
[bb]The Francis Crick Institute and Department of Medicine, Faculty of Medicine, Imperial College London, London, United Kingdom
[cc]CIRAD, UMR AGAP, Capesterre Belle Eau, Guadeloupe, France
[dd]AGAP, Université Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France
[ee]Imperial College London, Silwood Park Campus, Ascot, Berkshire, United Kingdom
[ff]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA
[gg]Integrated Research Facility at Fort Detrick, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Fort Detrick, Frederick, Maryland, USA

Accepted manuscript posted online 4 April 2018

Citation Krupovic M, Blomberg J, Coffin JM, Dasgupta I, Fan H, Geering AD, Gifford R, Harrach B, Hull R, Johnson W, Kreuze JF, Lindemann D, Llorens C, Lockhart B, Mayer J, Muller E, Olszewski NE, Pappu HR, Pooggin MM, Richert-Pöggeler KR, Sabanadzovic S, Sanfaçon H, Schoelz JE, Seal S, Stavolone L, Stoye JP, Teycheney P-Y, Tristem M, Koonin EV, Kuhn JH. 2018. *Orthervirales*: new virus order unifying five families of reverse-transcribing viruses. J Virol 92:e00515-18. https://doi.org/10.1128/JVI.00515-18.

Editor Rozanne M. Sandri-Goldin, University of California, Irvine

Address correspondence to Mart Krupovic, krupovic@pasteur.fr.

KEYWORDS *Belpaoviridae*, *Caulimoviridae*, *Hepadnaviridae*, International Committee on Taxonomy of Viruses (ICTV), *Metaviridae*, *Pseudoviridae*, *Retroviridae*, retroviruses, Ty3/Gypsy and Ty1/Copia LTR retrotransposons, virus classification, virus evolution

Reverse-transcribing viruses, which synthesize a copy of genomic DNA from an RNA template, are widespread in animals, plants, algae, and fungi (1, 2). This broad distribution suggests the ancient origin(s) of these viruses, possibly concomitant with the emergence of eukaryotes (3). Reverse-transcribing viruses include prominent human pathogens, such as human immunodeficiency viruses 1 and 2 (HIV-1/2) and hepatitis B virus, as well as plant pathogens that cause considerable economic losses (4).

The International Committee on Taxonomy of Viruses (ICTV) traditionally classified reverse-transcribing viruses into five families: *Caulimoviridae*, *Hepadnaviridae*, *Metaviridae*, *Pseudoviridae*, and *Retroviridae* (5). In 2018, the ICTV recognized an additional family, *Belpaoviridae*, which contains the genus *Semotivirus* (previously included in the *Metaviridae* family [6]). The infection cycles, nucleic acid types, genome organizations, and virion morphologies of these viruses are very diverse. Indeed, reverse-transcribing viruses are distributed between two Baltimore classes of viruses. Belpaoviruses, metaviruses, pseudoviruses (better known as Bel/Pao, Ty3/Gypsy, and Ty1/Copia retrotransposons, respectively [1, 7]), and retroviruses typically encapsidate single-stranded RNA (ssRNA) genomes (Table 1) and frequently integrate into the host genomes as part of their replication cycles (Baltimore class VI). In contrast, members of the families *Caulimoviridae* and *Hepadnaviridae*, often referred to as pararetroviruses (8), package circular double-stranded DNA (dsDNA) genomes and do not actively integrate into host chromosomes (Baltimore class VII). However, capture of pararetroviral DNA in host genomes, presumably by illegitimate recombination, is commonplace, particularly in plants, giving rise to the corresponding endogenous elements (9, 10).

Mechanistic studies of the replication cycles of reverse-transcribing viruses of different families have revealed many similarities, which have been reinforced by comparative genomics of the viral reverse transcriptases (RTs), the hallmark enzymes encoded by all reverse-transcribing viruses. Indeed, phylogenetic analyses support the monophyly of all viral RTs, excluding those encoded by nonviral retroelements from both eukaryotes and prokaryotes (11, 12). In addition to sharing RT phylogeny, belpaoviruses, caulimoviruses, metaviruses, pseudoviruses, and retroviruses share several conserved features that hepadnaviruses lack (Table 1). In particular, the polymerase (Pol) polyproteins of belpaoviruses, metaviruses, pseudoviruses, and retroviruses possess similar domain architectures. These Pol polyproteins contain an aspartate protease, which is responsible for the processing of viral polyproteins, and an integrase of the DDE recombinase superfamily. The genomes of these viruses also share long terminal repeats (LTRs) (13). Within certain clades, Pol polyproteins of retroviruses and metaviruses share additional features, such as a dUTPase domain (14–16) and the GPY/F subdomain of the integrase (17, 18). Caulimoviruses also possess a homologous aspartate protease domain in their Pol polyprotein (19) but lack an integrase and LTRs. However, RT-based phylogenies consistently classify these plant-infecting viruses as a sister clade of the metaviruses (Fig. 1). This phylogenetic position suggests that among pararetroviruses, encapsidation of a DNA genome is a homoplasious trait and, therefore, is not a reliable criterion for classification. The basal branches of the RT tree are not resolved and are presented as a multifurcation in Fig. 1. This topology is at least compatible with placing the *Hepadnaviridae* clade outside the viral group that includes belpaoviruses, caulimoviruses, metaviruses, pseudoviruses, and retroviruses.

Belpaoviruses, caulimoviruses, metaviruses, pseudoviruses, and retroviruses share not only homologous proteins involved in genome replication and polyprotein processing but also the two principal protein components of the virions, namely, the capsid and nucleocapsid proteins/domains (20–22). However, the nucleocapsid domain appears to be absent in spumaretroviruses (family *Retroviridae*) (Table 1). In contrast,

TABLE 1 Features shared by reverse-transcribing viruses[a]

Family or subfamily	Presence of Pol:			Presence of Gag:		Presence of LTR	Priming mechanism	Genome type
	RT-RH	Protease	Integrase	CA/CP	NC			
Retroviridae								
Orthoretrovirinae	+	+	+	+	+	+	tRNA	ssRNA
Spumaretrovirinae	+	+	+	+	–	+	tRNA	ssRNA/dsDNA[b]
Metaviridae	+	+	+	+	+	+	tRNA	ssRNA
Pseudoviridae	+	+	+	+	+	+	tRNA	ssRNA
Belpaoviridae	+	+	+	+	+	+	tRNA	ssRNA
Caulimoviridae	+	+	–	+	+	–[c]	tRNA	dsDNA
Hepadnaviridae	+	–	–	–	–	–[d]	TP	dsDNA

[a]Abbreviations: CA/CP, capsid protein; Gag, group-specific antigen; LTR, long terminal repeat; NC, nucleocapsid protein; RH, RNase H; RT, reverse transcriptase; Pol, polymerase polyprotein; TP, terminal protein.
[b]Members of the subfamily *Spumaretrovirinae* contain both ssRNA and dsDNA in extracellular particles, and reverse transcription occurs during virus assembly and disassembly.
[c]In the genus *Petuvirus* (*Caulimoviridae*), an inactivated integrase-like domain and quasi (long)-terminal repeats have been identified (28, 29), suggesting that certain ancestral elements have been lost during the evolution of caulimoviruses.
[d]Upstream of the capsid protein gene, hepadnavirus genomes contain a sequence showing similarity to the U5 region of the retroviral LTR (30).

hepadnaviruses encode an unrelated capsid protein (23). These findings suggest that belpaoviruses, caulimoviruses, metaviruses, pseudoviruses, and retroviruses have evolved from a common viral ancestor, rather than from distinct capsid-less retrotransposons (20).

Finally, similarities between belpaoviruses, caulimoviruses, metaviruses, pseudoviruses, and retroviruses extend to the mechanism of replication priming. All these viruses utilize host tRNA molecules as primers for genome replication by reverse transcription (24), whereas hepadnaviruses use a specific protein priming mechanism mediated by the polymerase terminal protein domain (25).

Taken together, the common complement of proteins required for genome replication, polyprotein processing, and virion formation, the topology of the RT phylogenetic tree, and mechanistic similarities in genome replication present strong evidence that belpaoviruses, caulimoviruses, metaviruses, pseudoviruses, and retroviruses share an evolutionary origin. The hepadnaviruses, which typically (i) branch out at the base of the viral RT clade (Fig. 1), (ii) possess a unique capsid protein, and (iii) employ a distinct replication mechanism, appear to be more distantly related to all these virus families. In recognition of these relationships, the ICTV has recently regrouped the families *Belpaoviridae*, *Caulimoviridae*, *Metaviridae*, *Pseudoviridae*, and *Retroviridae* into an order, *Ortervirales* (*Orter*, an inversion of "retro," which was derived from reverse transcription; *-virales*, suffix for an order). This change in taxonomy acknowledges and formalizes the long-proposed evolutionary relationship among most groups of reverse-transcribing viruses (26). We note that although hepadnaviruses are not included in the order, they might be unified with other reverse-transcribing viruses at a higher taxonomic level in the future.

ACKNOWLEDGMENTS

This work was supported in part through Battelle Memorial Institute's prime contract with the U.S. National Institute of Allergy and Infectious Diseases (NIAID; contract no. HHSN272200700016I, J.H.K.). E.V.K. is supported by intramural funds from the U.S. Department of Health and Human Services (to the National Library of Medicine). S.S. acknowledges support from SRI Funds from Mississippi Agriculture and the Forestry Experiment Station of Mississippi State University. J.F.K. is supported by the CGIAR Research Program on Roots, Tubers and Bananas (RTB) and supported by CGIAR Fund Donors (http://www.cgiar.org/aboutus/our-funders/). M.K. is supported by l'Agence Nationale de la Recherche (France) project ENVIRA.

M. Krupovic, B. Harrach, S. Sabanadzovic, H. Sanfaçon, and J. H. Kuhn were members of the 2014–2017 International Committee on Taxonomy of Viruses (ICTV) Executive Committee. J. Blomberg, J. M. Coffin, H. Fan, R. Gifford, W. Johnson, D. Lindemann, J. Mayer, J. P. Stoye, and M. Tristem were members of the 2014–2017 ICTV *Retroviridae* Study Group. I.

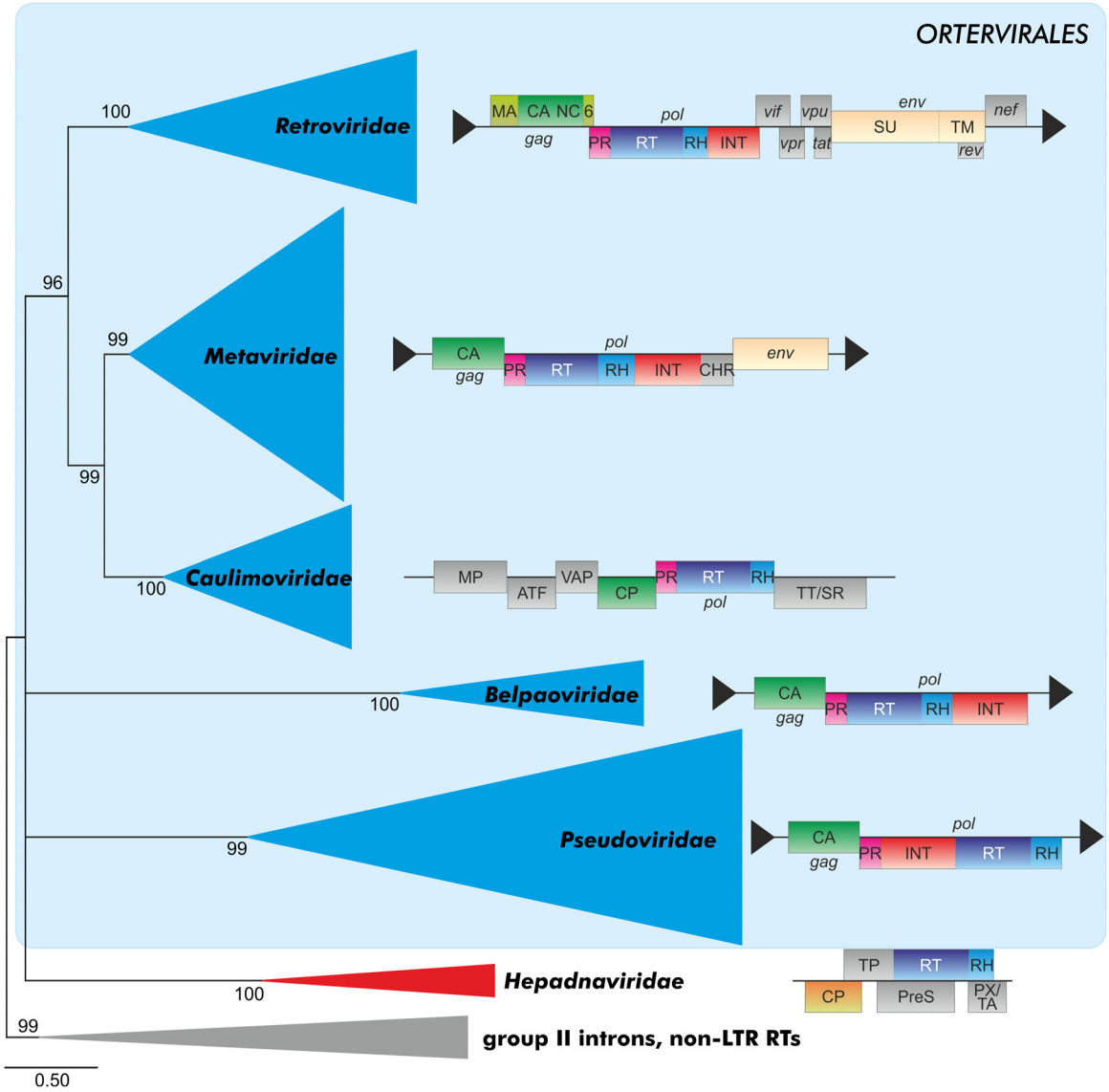

FIG 1 Maximum-likelihood phylogeny of viral reverse transcriptases. The tree includes sequences of 290 viruses belonging to all ICTV-recognized genera of reverse-transcribing viruses. The phylogeny was inferred using PhyML (27) with the LG+G+F substitution model and is rooted with sequences from nonviral retroelements (bacterial group II introns and eukaryotic LINE retroelements). Genomic organizations of selected representatives of reverse-transcribing viruses are shown next to the corresponding branches. Long terminal repeats are shown as black triangles. Note that members of the virus families display considerable variation in gene/domain content (5), which is not captured in this figure. Abbreviations: 6, 6-kDa protein; ATF, aphid transmission factor; CA/CP, capsid protein; CHR, chromodomain (present only in the integrase of particular clades of metaviruses of plants, fungi, and several vertebrates); *gag*, group-specific antigen; *env*, envelope genes; INT, integrase; LTR, long terminal repeat; MA, matrix protein; MP, movement protein; NC, nucleocapsid; *nef*, *tat*, *rev*, *vif*, *vpr*, and *vpu*, genes that express regulatory proteins via spliced mRNAs; P, polymerase; *pol*, polymerase gene; PR, protease; PreS, presurface protein (envelope); PX/TA, protein X/transcription activator; RH, RNase H; RT, reverse transcriptase; SU, surface glycoprotein; TM, transmembrane glycoprotein; TP, terminal protein domain; TT/SR, translation *trans*-activator/suppressor of RNA interference; VAP, virion-associated protein.

Dasgupta, A. D. Geering, R. Hull, J. F. Kreuze, B. Lockhart, E. Muller, N. Olszewski, H. R. Pappu, M. Pooggin, K. R. Richert-Pöggeler, J. E. Schoelz, S. Seal, L. Stavolone, and P.-Y. Teycheney were members of the 2014–2017 ICTV *Caulimoviridae* Study Group. R. Hull is retired from the John Innes Centre, Norwich, Norfolk, United Kingdom.

REFERENCES

1. Llorens C, Futami R, Covelli L, Dominguez-Escriba L, Viu JM, Tamarit D, Aguilar-Rodriguez J, Vicente-Ripolles M, Fuster G, Bernet GP, Maumus F, Munoz-Pomer A, Sempere JM, Latorre A, Moya A. 2011. The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. Nucleic Acids Res 39:D70–D74. https://doi.org/10.1093/nar/gkq1061.
2. Ahlquist P. 2006. Parallels among positive-strand RNA viruses, reverse-

transcribing viruses and double-stranded RNA viruses. Nat Rev Microbiol 4:371–382. https://doi.org/10.1038/nrmicro1389.
3. Koonin EV, Dolja VV, Krupovic M. 2015. Origins and evolution of viruses of eukaryotes: the ultimate modularity. Virology 479-480:2–25. https://doi.org/10.1016/j.virol.2015.02.039.
4. Geering ADW. 2014. *Caulimoviridae* (plant pararetroviruses). *In* eLS. John Wiley & Sons, Ltd, Chichester, United Kingdom. https://doi.org/10.1002/9780470015902.a0000746.pub3.
5. King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ. 2011. Virus taxonomy. Ninth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, CA.
6. Eickbush T, Boeke JD, Sandmeyer SB, Voytas DF. 2011. *Metaviridae*, p 457–466. *In* King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (ed), Virus taxonomy: classification and nomenclature of viruses. Ninth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, CA.
7. Arkhipova IR. 2017. Using bioinformatic and phylogenetic approaches to classify transposable elements and understand their complex evolutionary histories. Mob DNA 8:19. https://doi.org/10.1186/s13100-017-0103-2.
8. Hull R, Will H. 1989. Molecular biology of viral and nonviral retroelements. Trends Genet 5:357–359. https://doi.org/10.1016/0168-9525(89)90151-0.
9. Feschotte C, Gilbert C. 2012. Endogenous viruses: insights into viral evolution and impact on host biology. Nat Rev Genet 13:283–296. https://doi.org/10.1038/nrg3199.
10. Diop SI, Geering ADW, Alfama-Depauw F, Loaec M, Teycheney PY, Maumus F. 2018. Tracheophyte genomes keep track of the deep evolution of the *Caulimoviridae*. Sci Rep 8:572. https://doi.org/10.1038/s41598-017-16399-x.
11. Gladyshev EA, Arkhipova IR. 2011. A widespread class of reverse transcriptase-related cellular genes. Proc Natl Acad Sci U S A 108:20311–20316. https://doi.org/10.1073/pnas.1100266108.
12. Xiong Y, Eickbush TH. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J 9:3353–3362.
13. Benachenhou F, Sperber GO, Bongcam-Rudloff E, Andersson G, Boeke JD, Blomberg J. 2013. Conserved structure and inferred evolutionary history of long terminal repeats (LTRs). Mob DNA 4:5. https://doi.org/10.1186/1759-8753-4-5.
14. Mayer J, Meese EU. 2003. Presence of dUTPase in the various human endogenous retrovirus K (HERV-K) families. J Mol Evol 57:642–649. https://doi.org/10.1007/s00239-003-2514-6.
15. Rodriguez F, Kenefick AW, Arkhipova IR. 2017. LTR-retrotransposons from bdelloid rotifers capture additional ORFs shared between highly diverse retroelement types. Viruses 9:E78. https://doi.org/10.3390/v9040078.
16. Novikova OS, Blinov AG. 2008. dUTPase-containing metaviridae LTR retrotransposons from the genome of *Phanerochaete chrysosporium* (Fungi: Basidiomycota). Dokl Biochem Biophys 420:146–149. https://doi.org/10.1134/S1607672908030137.
17. Jern P, Sperber GO, Blomberg J. 2005. Use of endogenous retroviral sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy. Retrovirology 2:50. https://doi.org/10.1186/1742-4690-2-50.
18. Malik HS, Eickbush TH. 1999. Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons. J Virol 73:5186–5190.
19. Marmey P, Rojas-Mendoza A, de Kochko A, Beachy RN, Fauquet CM. 2005. Characterization of the protease domain of Rice tungro bacilliform virus responsible for the processing of the capsid protein from the polyprotein. Virol J 2:33. https://doi.org/10.1186/1743-422X-2-33.
20. Krupovic M, Koonin EV. 2017. Homologous capsid proteins testify to the common ancestry of retroviruses, caulimoviruses, pseudoviruses, and metaviruses. J Virol 91:e00210-17. https://doi.org/10.1128/JVI.00210-17.
21. Vo JN, Campbell PR, Mahfuz NN, Ramli R, Pagendam D, Barnard R, Geering AD. 2016. Characterization of the banana streak virus capsid protein and mapping of the immunodominant continuous B-cell epitopes to the surface-exposed N terminus. J Gen Virol 97:3446–3457. https://doi.org/10.1099/jgv.0.000643.
22. Sandmeyer S, Patterson K, Bilanchone V. 2015. Ty3, a position-specific retrotransposon in budding yeast. Microbiol Spectr 3:MDNA3-0057-2014. https://doi.org/10.1128/microbiolspec.MDNA3-0057-2014.
23. Steven AC, Conway JF, Cheng N, Watts NR, Belnap DM, Harris A, Stahl SJ, Wingfield PT. 2005. Structure, assembly, and antigenicity of hepatitis B virus capsid proteins. Adv Virus Res 64:125–164. https://doi.org/10.1016/S0065-3527(05)64005-5.
24. Menendez-Arias L, Sebastian-Martin A, Alvarez M. 2017. Viral reverse transcriptases. Virus Res 234:153–176. https://doi.org/10.1016/j.virusres.2016.12.019.
25. Nassal M. 2008. Hepatitis B viruses: reverse transcription a different way. Virus Res 134:235–249. https://doi.org/10.1016/j.virusres.2007.12.024.
26. Hull R. 2001. Classifying reverse transcribing elements: a proposal and a challenge to the ICTV. International Committee on Taxonomy of Viruses. Arch Virol 146:2255–2261.
27. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–321. https://doi.org/10.1093/sysbio/syq010.
28. Richert-Pöggeler KR, Noreen F, Schwarzacher T, Harper G, Hohn T. 2003. Induction of infectious petunia vein clearing (pararetro) virus from endogenous provirus in petunia. EMBO J 22:4836–4845. https://doi.org/10.1093/emboj/cdg443.
29. Richert-Pöggeler KR, Shepherd RJ. 1997. Petunia vein-clearing virus: a plant pararetrovirus with the core sequences for an integrase function. Virology 236:137–146. https://doi.org/10.1006/viro.1997.8712.
30. Miller RH, Robinson WS. 1986. Common evolutionary origin of hepatitis B virus and retroviruses. Proc Natl Acad Sci U S A 83:2531–2535.